R E P O R T R E S U M E S

ED 010 881 AL 000 103

A SCHEMA FOR SENTENCE COORDINATION.

BY- SCHANE, SANFORD A.
MITRE CORP., BEDFORD, MASS.

REPORT NUMBER MTP-10 PUB DATE APR 66
EDRS PRICE MF-$0.18 HC-$2.60 65P.

DESCRIPTORS- *COMPUTATIONAL LINGUISTICS, *TRANSFORMATION THEORY (LANGUAGE), TRANSFORMATIONS (LANGUAGE), *ENGLISH, PHRASE STRUCTURE, *LINGUISTIC THEORY, COMPUTER PROGRAMS, MITRE GRAMMAR

A COORDINATION SCHEMA IS DESCRIBED WHICH HAS BEEN INTRODUCED IN THE "MITRE" GRAMMAR. THIS SCHEMA DEALS WITH THE TYPE OF COORDINATION WHERE A SINGLE SENTENCE OF TWO OR MORE CONSTITUENTS ALLOWS FOR A PARAPHRASE OF TWO OR MORE SEPARATE SENTENCES WHICH ARE IDENTICAL IN EVERY RESPECT EXCEPT FOR THE CONJOINABLE MEMBERS. THE NEW APPROACH REPLACES EARLIER APPROACHES IN GENERATIVE GRAMMAR WHICH HANDLED COORDINATE CONSTRUCTIONS BY USING GENERALIZED TRANSFORMATIONS, AND LATER EMBEDDING TRANSFORMATIONS. THE REPORT CONTAINS TWO SECTIONS. THE FIRST IS THEORETICAL AND DESCRIBES THE NEW SCHEMA WHICH CONSISTS OF TWO SETS OF RULES--(1) THE PRIMARY CONJUNCTION RULES WHICH DERIVE A SINGLE SENTENCE WITH CONJOINED CONSTITUENTS FROM TWO OR MORE COORDINATE SENTENCES, AND (2) THE SECONDARY CONJUNCTION RULES WHICH OPERATE ON THESE DERIVED COORDINATE STRUCTURES, CONVERTING THEM TO RELATED VARIANT FORMS. THIS SCHEMA ENTAILS A REEXAMINATION OF SOME OF THE PHRASE STRUCTURE RULES. THE SECOND PART OF THE REPORT DISCUSSES THE IMPLEMENTATION OF THIS SCHEMA. (KL)

ED010881

NOV 9 1966

APRIL 1966/MTP-10

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

a schema for sentence coordination

SANFORD A. SCHANE

Information Sciences Department
The MITRE Corporation

AL 000 103

AL 000 103

INFORMATION SYSTEM LANGUAGE STUDIES NUMBER TEN...

INFORMATION SYSTEM LANGUAGE STUDIES NUMBER TEN MTP-10

A SCHEMA FOR SENTENCE COORDINATION

by

Sanford A. Schane
Information Sciences Department
(with an Appendix by Joyce Friedman)

April 1966

THE MITRE CORPORATION
BOX 208
BEDFORD, MASSACHUSETTS

This document has been released for public dissemination.

Contract No. AF 19(628)-5165
MITRE Project No. 7020

FOREWORD

I should like to express my gratitude to the linguists who were at The MITRE Corporation during the summer of 1965 and with whom I was able to discuss the problems which led to the writing of this paper. In particular, I am indebted to Paul Chapin, who offered valuable criticism on an earlier version of the paper. His insightful ideas have been incorporated into the study presented here.

Stanford A. Schane
University of California at
San Diego

PRECEEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

		Page
0.	INTRODUCTION	1
1.	TYPES OF COORDINATION	3
2.	DERIVATION	5
3.	TYPES OF CONJOINED CONSTRUCTIONS	9
4.	SCHEMA FOR COORDINATION	13
5.	AMBIGUITIES IN COORDINATION	16
6.	EXAMPLES OF PRIMARY CONJUNCTION RULES	22
7.	PHRASE STRUCTURE RULES	27
8.	INTERROGATIVES	30
9.	THE OPTIONALITY OF PRIMARY CONJUNCTION	35
10.	SUBJECT-PREDICATE AGREEMENT	36
11.	AN ADDITIONAL CONSTRAINT ON PRIMARY CONJUNCTION	38
12.	CONJUNCTION AND THE ENGLISH PREPROCESSOR	39
APPENDIX A	(The Test Grammar)	43
	Phrase Structure Rules	43
	Transformations	45
APPENDIX B	(Programming of Conjunction)	56
	1. Strategy	56
	2. Primary Conjunction Transformation	56
	3. Conjunction Subroutine	57
	4. Instructions for Conjunction	57
APPENDIX C	(References)	60

0. INTRODUCTION

Within a generative grammar coordinate constructions are handled awkwardly if they are derived by means of transformational rules. Yet that many types of coordinate constructions should originate from conjoined full sentences, which are subsequently reduced in some way, seems to be well motivated. This is, furthermore, by no means a recent or novel observation.

In the earlier works on generative grammar one finds two types of transformational rules: singulary and generalized. The singulary transformations operate on a single sentence, adding, deleting, or permuting constituents or substituting one constituent for another. Generalized transformations operate on a pair of sentences--a matrix sentence and a constituent sentence--embedding the latter into the former. A later theoretical development was the notion that all recursion should be through the base or phrase structure grammar, by the reintroduction of the initial symbol #S# in the phrase structure expansions of some grammatical symbols. The new notion of "embedding transformation" - any transformation which applies across a sentence boundary - supplanted that of the generalized transformation. Thus, embedding transformations account for complex sentences, various types of complement constructions, and complex adverbial and adjectival modifiers. Since certain coordinate constructions are to be derived from two or more full sentences, it might seem that embedding transformations are the appropriate vehicle for reducing conjoined sentences to a single sentence containing coordinate constituents.

It was this approach using embedding transformations which was originally adopted for treating coordination within the MITRE grammar. This allowed-although somewhat clumsily--for a restricted number of coordinate constructions. In some cases the rules also permitted the generation of unacceptable sentences, whereas at the other extreme, there were perfectly acceptable coordinate constructions which could not be handled at all. From an examination of the various coordination

rules, it was furthermore evident that there were generalizations that were not being captured. Consequently, coordination was re-examined during the summer of 1965, and it was proposed that coordination be handled primarily by a special schema, non-transformational in nature. Subsequent computer trials have indicated the feasibility of this proposed approach.

This report is divided into two sections. The first part is theoretical. Certain observations concerning coordinate structures are made. The theoretical objections to treating coordination by means of generalized transformations are set forth and a different schema is proposed to replace such transformations. This schema is partly transformational and partly non-transformational in nature. The non-transformational part--what we call the primary conjunction rules--derives a single sentence with conjoined constituents from two or more coordinate sentences. A set of singulary transformations--the secondary conjunction rules--may then operate on these derived coordinate structures, converting them to related variant forms. This schema for coordination forces us to re-examine some of the most elementary phrase structure rules, particularly those which have to do with the first occurrence of the auxiliary.

The second part of the report (with the appendices) deals with the implementation of the coordination schema within the MITRE grammar. The problems dealt with here are less general in nature and primarily involve accommodating the schema to the constraints which have been imposed by the programming. The rules for primary conjunction and some of the secondary conjunction rules are given in the MITRE format. In Appendix B Joyce Friedman has outlined the actual programming of conjunction. Due to the specialized nature of the second half of this report, the reader should be familiar with the format of the MITRE grammar and with the notational conventions used in writing rules (see reference list in Appendix C). However, the first part, which is concerned with theoretical aspects, does not require familiarity with the MITRE grammar and hence can be read independently.

1. TYPES OF COORDINATION

This report deals with the type of coordination where a single sentence containing two (or more) conjoined constituents allows for a paraphrase of two (or more) separate sentences which are identical in every respect except for the conjoinable members. Thus, (1) but not (2) falls within the realm of the discussion:

(1) John and Mary are here.
  (a) John is here.
  (b) Mary is here.

(2) John and Mary are a happy couple.
  (a) *John is a happy couple.
  (b) *Mary is a happy couple.

The observation that sentences such as (1) are related to the corresponding independent sentences (1a) and (1b) has been noted by traditional grammarians. Thus, Curme states: "The members [of a compound sentence] are connected by coordinating conjunctions: 'John is in the garden working and Mary is sitting at the window reading.' The members of a compound sentence, however, are not always thus complete, each with subject and finite verb, for a natural feeling for the economy of time and effort prompts us, wherever it is possible, to contract by employing a common verb for all members, so that the conjunctions connect only parts of like rank: not 'John is writing and Mary is writing,' but 'John and Mary are writing,' or 'John and Mary are both writing,' or 'Both John and Mary are writing' (George O. Curme, Syntax Boston, 1931, p. 161).

The distinction between sentence types (1) and (2) is also noted by Curme: "Sentences containing these conjunctions, however, are often not an abridgement of two or more sentences, but a simple sentence with elements of equal rank, connected by a conjunction: "'The King and Queen are an amiable pair.' 'She mixed oil and water together'" (p. 162). Sentences such as (2) do not admit of the 'both ... and' construction: *'Both John and Mary are a happy couple.'

*'She mixed both wine and oil together.' Therefore, the possibility of a 'both...and' paraphrase can be used as a criterion for determining the set of sentences with coordinate constructions which are derivable from two (or more) conjoined sentences.[1]

Some sentences are ambiguous in that they can be interpreted either as type (1) or as type (2). A sentence such as (3) therefore, can have the readings of (3a) or (3b).

(3) John and Helen are married.

(a) John and Helen are each married.

(b) John and Helen are married to each other.

The interpretation (3a) merely reports on the marital status of the individuals (compare 'John and Bill are married.'). It is neutral as to whether John and Helen are husband and wife or whether each is married to someone else. Therefore, whenever (3) has the interpretation (3a), then (3) like (1) can be related to two independent sentences. However, when (3) has the interpretation given in (3b), it cannot be so paraphrased.

If sentence (3) occurs with the 'both...and' construction it is no longer ambiguous.

(4) Both John and Helen are married.

(4) uniquely has the interpretation given in (3a). This corroborates the observation that only conjoined elements of type (1) allow the 'both... and' construction.

Ambiguities such as (3) can arise only when the coordinating conjunction is and. With the conjunction or, type (2) sentences are completely excluded. Thus, (5), unlike (3), has only one interpretation - that of (3a) - whereas (6) is unacceptable.

[1]The occurrence of 'both...and' is probably limited to two conjoined constituents. Note, however, that 'both...and' does not occur with full sentence coordination (i.e., compound sentences); e.g., *Both John is a fool and Mary is the darling of society.

(5) John or Helen is married.

(6) *John or Helen is a happy couple.

Therefore, a sentence in which the constituents are conjoined by or can always be related to two independent sentences.[2]

This report will not investigate further coordination types (2) and (3). It will concentrate entirely on type (1) coordination, where the conjoined constituents are derived from two (or more) independent sentences.

2. DERIVATION

Since coordinate constructions are to be derived from two or more independent sentences, a generative grammar composed of rewrite rules must provide the means for generating any number of conjoined sentences. For illustrative purposes a much simplified set of phrase structure rules is given in (7)

(7) (a) SS → #S#

(b) $S \rightarrow \left\{ \begin{matrix} \#S\# \ (\{ \begin{matrix} AND \\ OR \end{matrix} \} \ \#S\#)^* \\ NP \quad VP \end{matrix} \right\}$

(c) VP → BE PRED

(d) PRED → here

(e) $NP \rightarrow \left\{ \begin{matrix} John \\ Mary \\ Charlie \end{matrix} \right\}$

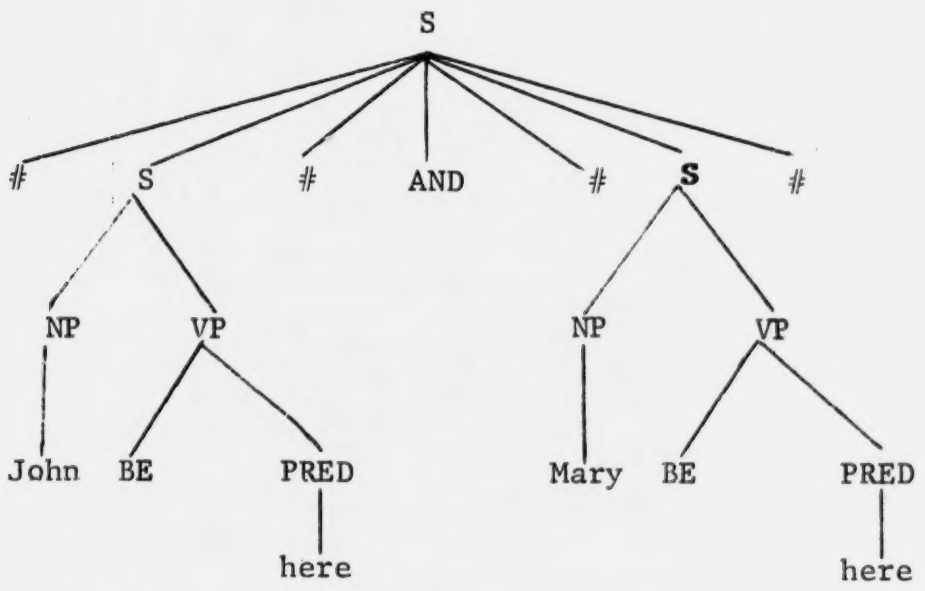
[2]When the conjunction is and and the conjoined constituents are nominal expressions a third type of coordination is possible, e.g., 'Bread and butter is served at most meals.' These sentences, like type (2), do not admit of the 'both...and' construction. However, they differ from type (2) in that when the conjoined nominal occurs in subject position the verb is in the singular. Thus, the nominal expression 'bread and butter' can occur in all three types:

Type 1: (Both) bread and butter are sold in most stores.
Type 2: Bread and butter are eaten together.
Type 3: Bread and butter is served at most meals.

The nominals which can occur as type (3) are highly restricted e.g., 'ham and eggs, pencil and paper.' The conjoined members have a fixed ordering; e.g., not butter and bread.

In rule (7b) the asterisk indicates that one or more occurrences of $\left\{ \begin{matrix} \text{AND} \\ \text{OR} \end{matrix} \right\}$ #S# may be selected. Rule (7b) thus permits one to expand the initial symbol S into a string of one or more S's such that all S's after the first are preceded by one of the coordinating conjunctions (and, or). The rules given in (7) allow one to derive conjoined sentences such as (8):

(8)



Rule (7b) is recursive in that it can be reapplied to any of the S's resulting from it. Structural ambiguities should be possible as a result of this recursiveness. Thus, sentence (9) has the three possible interpretations[3] shown in (10):

[3]Sentence (9) can be disambiguated by means of appropriate intonation contours. See L. R. Gleitman, "Coordinating Conjunctions in English," Language 41.2 (1965). Gleitman has suggested that whenever a form such as (9) has the branching illustrated in (11a) (i.e., is derived from (11a)), then all coordinate conjunctions except the last are obligatorily deleted; e.g., John and Mary and Charlie are here → John, Mary, and Charlie are here.

(9) John and Mary and Charlie are here.

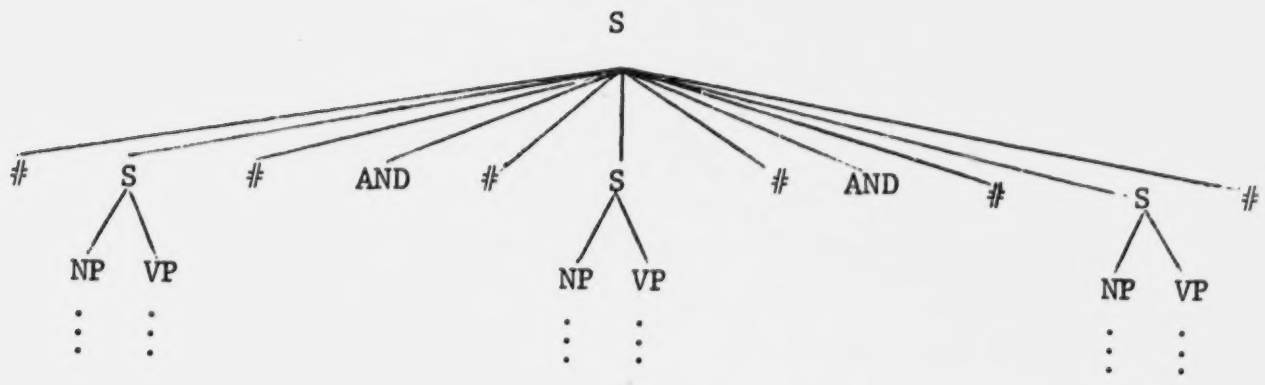
(10) (a) (John, Mary, and Charlie) are here.

(b) Both (John and Mary) and Charlie are here.

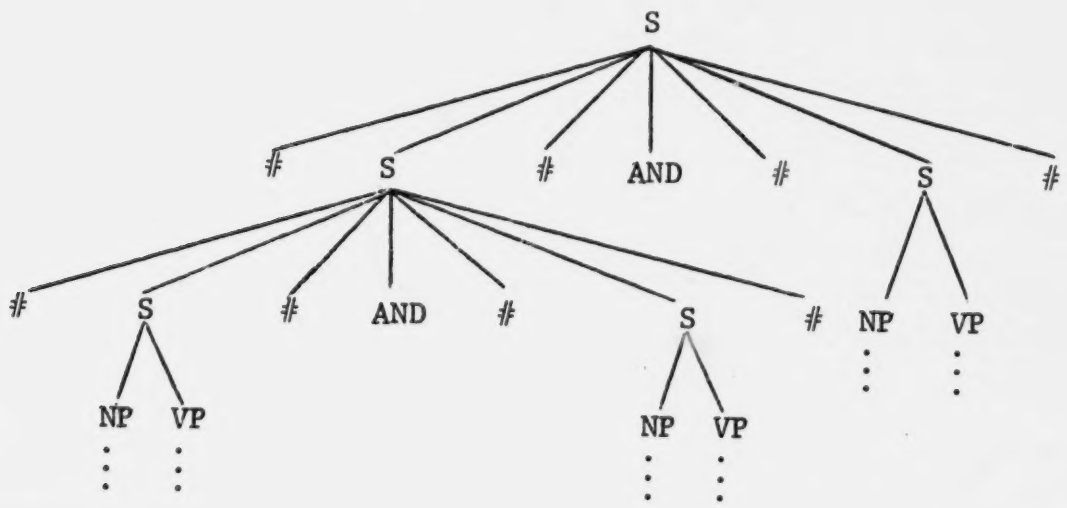
(c) Both John and (Mary and Charlie) are here.

(9) is in all cases derived from three independent sentences. The ambiguity is due to the number of conjoined sentences generated by rule (7b) and the reapplication of rule (7b). The base tree structures corresponding to (10a-c) respectively are shown in (11a-c).

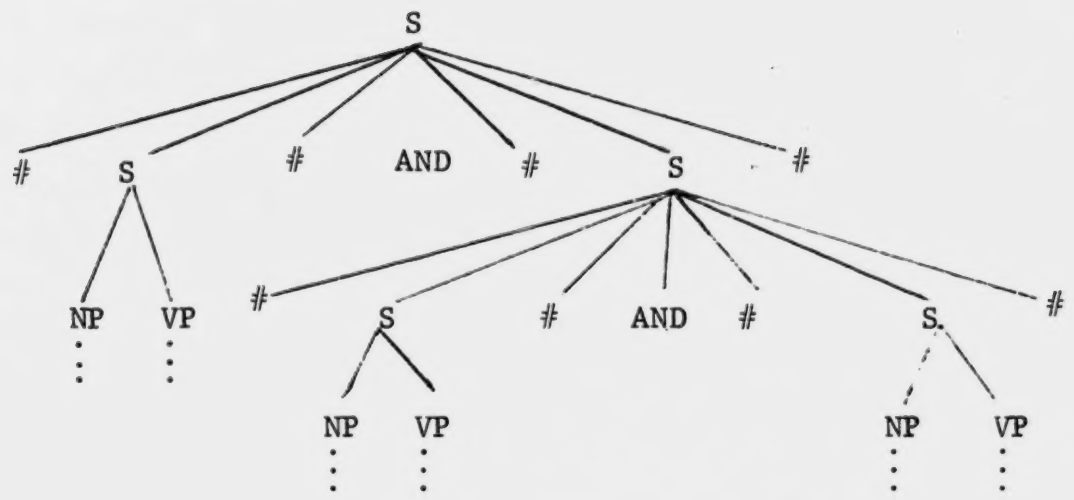
(11) (a)



(b)



(c)



In (11a) the initial symbol S has been expanded to a sequence of three conjoined S's; rule (7b) has been applied only once. In (11b) and (11c) the initial symbol S has been expanded to a sequence of two S's; rule (7b) has been reapplied to the first of these S's in (11b), and to the second in (11c).

Sentence (12) is like sentence (9) except that the first occurrence of and has been replaced by or.

(12) John or Mary and Charlie are here.

However, (12), unlike (9), is only two ways ambiguous.

(13) (a) Both (John and Mary) and Charlie are here.

(b) Either John or (Mary and Charlie) are here.

(12) can only have the structures shown in (11b) and (11c). (11a) is possible only if all of the coordinating conjunctions are identical. Therefore, rule (7b), which allows for the generation of any number of conjoined sentences, has to be constrained so that only the correct set of conjoined sentences will be obtained. Consideration of the examples shows that the proper constraint is that all of the conjunctions at any one level of compounding, that is, conjoining S's all dominated by the same node, must be the same. This is actually a constraint on the interpretation of the asterisk notation, and seems quite reasonable.

Rule (7b) will account for compound sentences such as (14).

(14) John goes to college and Mary works in a department store.

However, it was noted that when certain constituents of conjoined sentences are identical, the independent sentences are reduced: "not 'John is writing and Mary is writing ,' but 'John and Mary are writing.'" It remains to determine which constituents are to be reduced and how to state formally this reduction.

3. TYPES OF CONJOINED CONSTRUCTIONS

So far all examples illustrating coordinate constituents have been limited to conjoined nominals functioning as subject of sentence. Other conjoined constituents are possible. A few of these are noted in (15).

(15) (a) The plane landed at 1200 hours and will leave at 1500 hours. (predicate coordination)

(b) John caught and ate fish. (verb coordination)

(c) John saw Mary and Helen. (object coordination)

(d) The cat moved swiftly and quietly. (manner adverbial coordination)

(e) The plane leaves for Detroit at 8:00 am and at 9:30 pm. (time adverbial coordination)

(f) I shall be going to London and Paris. (place adverbial coordination)

(g) Most receptionists are young and beautiful. (adjective coordination)

(h) All her friends are doctors or lawyers. (noun predicate coordination)

The underlying base tree for the sentence 'John and Mary are here' was given in (8), where the phrase structure component of the grammar has generated the two conjoined sentences: 'John BE here AND Mary BE here.' In order to reduce these two sentences to a single sentence with a conjoined noun phrase, one could make use of the following transformational rule:

(16) $\#$ - NP_1 - VP_1 - $\#$ - AND - $\#$ - NP_2 - VP_2 - $\#$

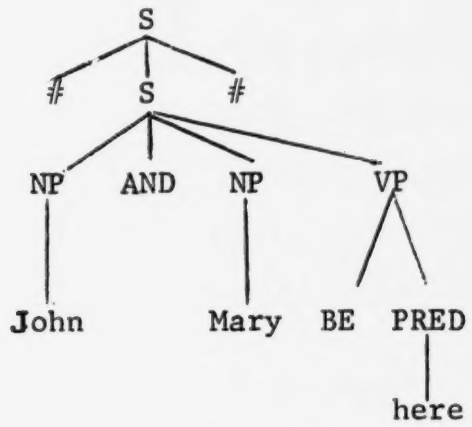
1 2 3 4 5 6 7 8 9 $\Rightarrow$
1 2+5+7 3 4 Ø Ø Ø Ø Ø

where $VP_1 = VP_2$

This rule states that AND and the NP of the second sentence are adjoined to the NP of the first sentence (i.e., in the derived tree structure AND NP_2 emanate from the same node as NP_1), provided that the VP are identical. The original AND - $\#$ - NP_2 - VP_2 - $\#$ is then deleted.

After applying (16) to base tree (8) the following derived tree is obtained:

(17)

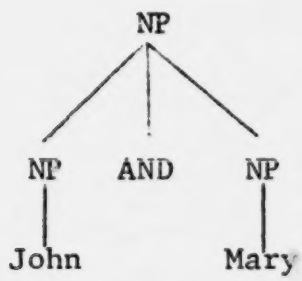


A subsequent rule could provide for the appropriate agreement between subject and verb.

In traditional grammars the sentence 'John and Mary are here' is described as a simple sentence having a compound subject. Except for subject verb agreement a compound subject functions within the sentence in the same manner as a simple subject, so that a transformation which applies to any subject NP should also apply to the conjoined NP's. For example, there is a transformation which converts any sentence of the type NP-BE-LOC to a corresponding sentence with introductory there, on condition that the NP dominates a non-definite determiner; e.g., A linguist is in the room ⇒ There is a linguist in the room. This transformation must also apply to a conjoined NP; e.g., A linguist and a programmer are in the room ⇒ There are a linguist and a programmer in the room.

This means that the whole nominal expression 'John and Mary' (NP AND NP) is to be considered as an NP. The desired structure is shown in (18).

(18)



The tree (17) does not exhibit this structure, since transformations such as (16) do not allow for new structure (grammatical nodes) to be created; i.e., the underlying tree has two NP nodes, whereas the desired surface tree would have three NP nodes. There is no way of deriving the additional dominating NP node. Without some ad hoc adjustment, transformational rules such as (16) are not capable of producing the desired derived structure.

Rule (16) as stated is an embedding transformation since it is defined as operating on two different S's, embedding part of one S into the other S. In this sense it is similar to such a transformation

as the one for relative clause embedding. However, this similarity is only apparent. In relative clause embedding there is a matrix (main) sentence which dominates a constituent (subordinate) sentence. However, in coordination one sentence is not contained within another, for conjoined sentences are of equal rank. Therefore, if coordination is treated as a type of embedding, there is no a priori way to determine which sentence is to be the embedding one and which is to be the embedded. In rule (16) the NP from the right-most sentence has been embedded into the sentence on the left. However, one could just as easily have stated the transformation so that the NP of the left-most sentence would be embedded into the sentence on the right. The arbitrariness of selecting a matrix sentence is of course increased with the number of conjoined sentences initially generated.

Rule (16) can only conjoin those nominals which function as subject of sentence. Additional transformational rules would be needed for conjoining predicate NP's, object NP's, adjectives, adverbials, etc. There would have to be as many rules as there are conjoinable grammatical categories and positions of occurrence. These rules would be similar in that they would all be conjoining constituents of like type. Yet the fact that there are separate rules means that this generalization is being missed. It would be desirable to replace all of the individual statements by a more general principle. Such a principle has been suggested by Chomsky in Syntactic Structures (Mouton, 1957, p. 36): "If S_1 and S_2 are grammatical sentences, and S_1 differs from S_2 only in that X appears in S_1 where Y appears in S_2 (i.e., S_1 = ..X.. and S_2 = ..Y..), and X and Y are constituents of the same type in S_1 and S_2, respectively, then S_3 is a sentence, where S_3 is the result of replacing X by X + and + Y in S_1 (i.e., S_3 = ..X + and +Y..)."

However, this principle is not a transformational rule.[4]

[4]Chomsky formulates a generalized transformation to correspond to this principle in Appendix II of Syntactic Structures, p. 113, Rule 22. However, the same comments which apply to the principle seem to apply to the proposed transformation.

Transformational rules are stated in terms of a proper analysis of a tree; i.e., the structural description of a transformation refers to a linear sequence of constituents. Thus, as can be seen from rule (16), the structural description refers to a sequence of constituents which are either non-terminal (NP, VP) or terminal (AND). It is these specified entities which are permuted, deleted, etc. In addition, although this is not reflected in rule (16), variables such as X, Y, Z (equivalent to ...) may be used to indicate a string of constituents of arbitrary length (X, Y, Z may also be null). The constituents encompassed by the variable are skipped over in the matching of a string to a structural description. Finally, certain structural conditions may be imposed. Thus, in (16) $VP_1 = VP_2$; the identity condition here demands that VP_1 and VP_2 both dominate the same subtrees (including lexical items). But in the principle, what is required is that "X and Y are constituents of the same type." That is to say, X and Y are variables across grammatical nodes, and an equality condition, to be distinguished from an identity condition, is defined on them. The subtrees which they dominate need not--indeed, may not--be identical; rather the grammatical nodes to be conjoined, whatever they may be, must be the same. It is precisely this condition which is beyond transformational power; and it is the incorporation of a version of this condition which distinguishes our coordination schema from a transformation. The coordinate constructions with which we are concerned are all derived from separate (underlying) sentences. What all these coordinate constructions have in common is that there is some part from each of the sentences which is identical and some part, dominated by constituents of like type, from each of the sentences which is different. It is the "different" constituents which are conjoinable.

4. SCHEMA FOR COORDINATION

If the identity and non-identity conditions on the independent sentences are appropriately met, then the tree composed of conjoined sentences will be replaced by a new tree composed of a single sentence having the non-identical constituents appropriately conjoined.

Furthermore, we shall demonstrate that the non-identical parts of the separate sentences must be restricted in such a way that they may differ by only the structure dominated by one grammatical node. The "principle for conjunction" can then be stated as follows:

(19) Two (or more) sentences can be conjoined into a single sentence if their total tree structure is identical except for the structure dominated by one grammatical node.

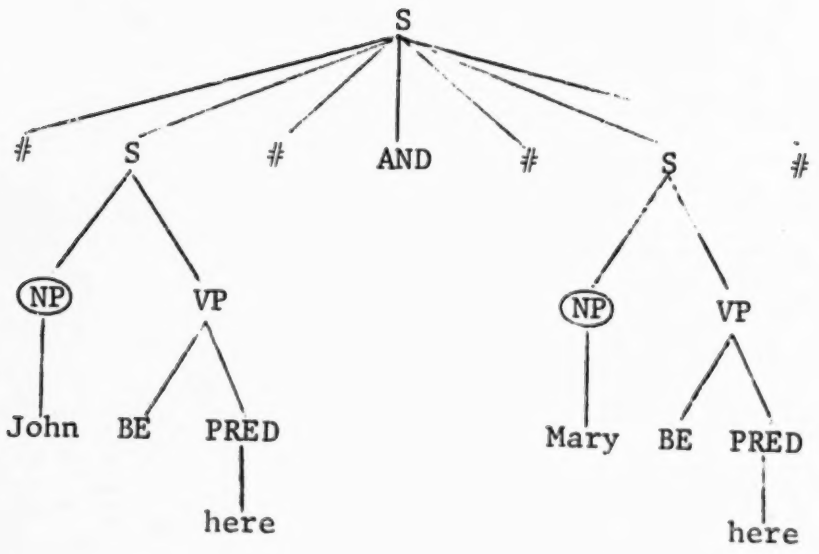
Thus, consider the tree of (8) where the underlying representation is: John BE here AND Mary BE here. If one compares the entire tree of each component sentence, it is seen that their tree structure is identical except for the NP node (i.e., it is the NP node which dominates a different lexical item in each case). Hence, these two sentences can be collapsed to a single sentence with a conjoined NP.

The restriction "except for the structure dominated by one grammatical node" will automatically exclude such questionable sentences as: 'John enjoyed and my friend liked the play' from 'John enjoyed the play and my friend liked the play.' (see Chomsky, 1957, pp. 35-36). Such a sentence differs by two grammatical nodes; i.e., the subject NP and the V; our principle then provides a basis for the exclusion of such deviant sentences.

If two (or more) sentences satisfy the conditions stated in the principle for conjunction, then these sentences can be reduced. The reduction results from the following four operations.

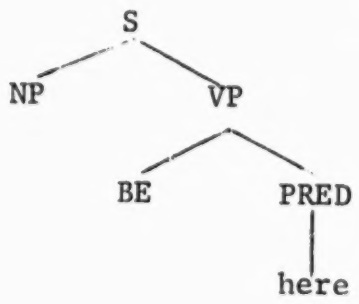
(i) The single grammatical node beneath which the tree structures differ is flagged.

(20)



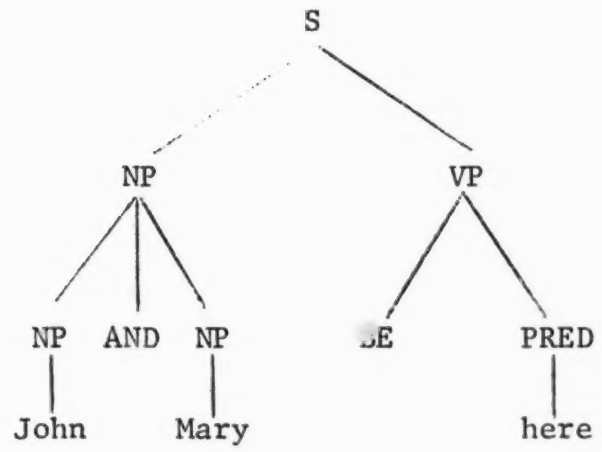
(ii) A single tree is reproduced which is equivalent to the tree of any one of the conjoined sentences down to the flagged constituent (i.e., all nodes beneath a flagged constituent are not reproduced.)

(21)



(iii) The flagged constituent--and all structure below it--of each of the conjoinable sentences (as well as the conjunction marker (and, or)) is attached to the flagged node of the new tree.

(22)



(iv) The old tree is erased.

Note that these operations must be performed simultaneously on any number of conjoined sentences.

5. AMBIGUITIES IN COORDINATION

The conjoined noun phrase 'the old men and women' is ambiguous,[5] the interpretations being:

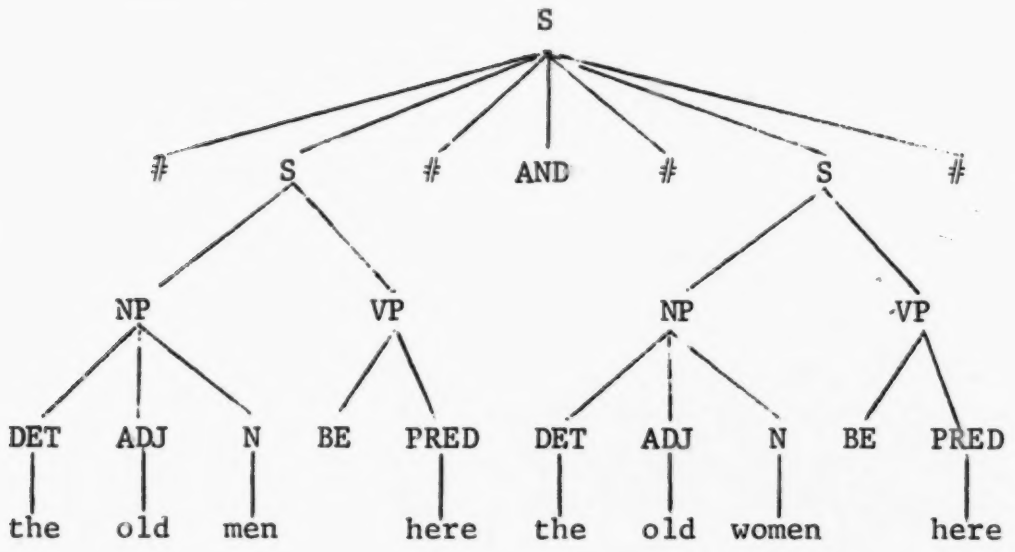
(23) (a) the old men and the old women
(b) the old men and the women

Thus, sentence (24) has the underlying tree structures shown in (25a) and (25b).

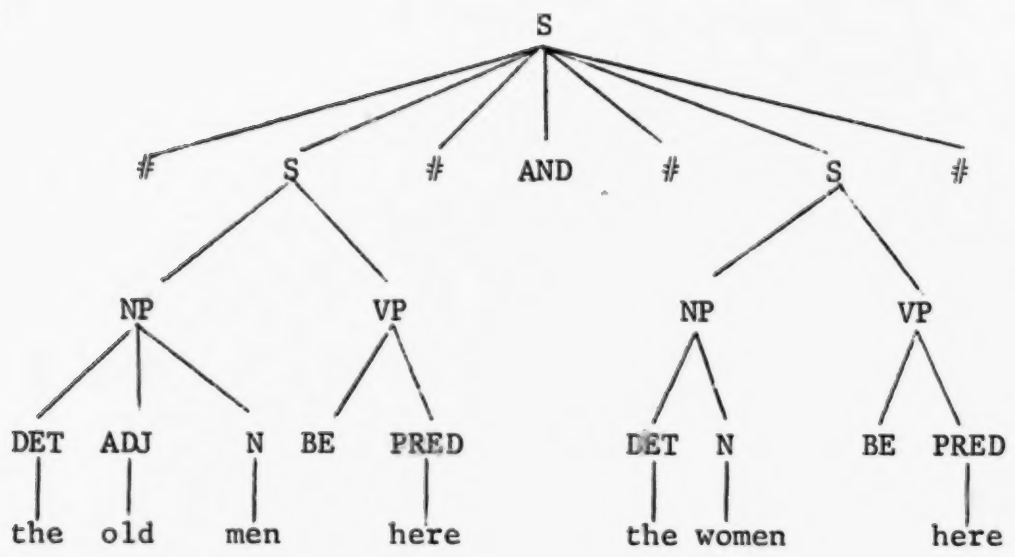
[5]See C. F. Hockett, "Two Models of Grammatical Description" Word, 1954, 10, 210-31.

(24) The old men and women are here.

(25) (a)

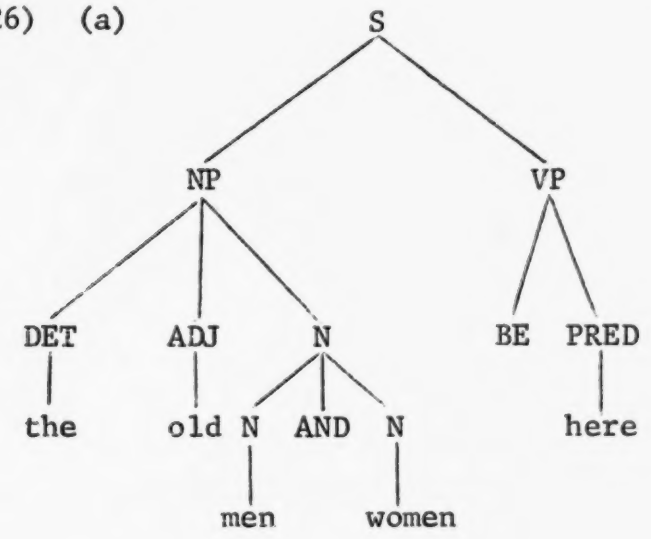


(b)

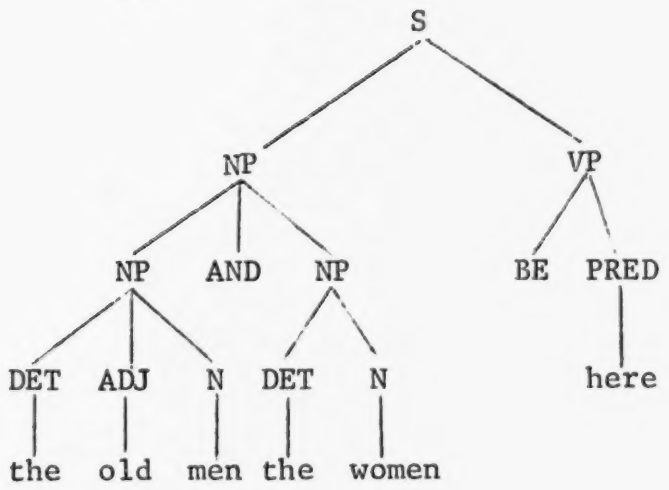


In 25(a) all grammatical nodes are identical except for N; in 25(b) all grammatical nodes are identical except for NP. After applying the conjunction operations the following trees are obtained:

(26) (a)



(b)



Thus, (25a) and (25b) do not both reduce to the ambiguous (24); although the principle for conjunction reduces (25a) to (24) (i.e., (26a)), it does not reduce (25b) to (24), but to (26b). The restriction concerning one grammatical node does not allow (25a) and (25b) to be reduced in the same way. In fact, further investigation of

conjoined constituents reveals that sentences which are reduced can often be reduced to more than one form. Thus, (27a), (27b), and (27c) are all perfectly acceptable reductions of (25a), and (28a) and (28b) are reductions of (25b).

(27) The old men BE here AND the old women BE here.

(a) The old men and the old women are here.
(b) The old men and old women are here.
(c) The old men and women are here.

(28) The old men BE here AND the women BE here.

(a) The old men and the women are here.
(b) The old men and women are here.

It has already been shown that the principle for conjunction in its present form allows one to derive only (27c) and (28a). Clearly one must be able to account for all of the derived sentences. This suggests that the restriction that all of the differing structure be dominated by one grammatical node must be more carefully defined if the principle for conjunction is to be valid.

In the above examples if one were to exclude coordination of N's and allow only NP coordination, one would obtain the fullest reduced forms; i.e., (27a) and (28a). The other possibilities could then be obtained by a subsequent set of deletion rules which would operate on derived conjoined strings, deleting repeated elements if certain conditions are met.

The following two deletion rules would then permit one to derive (27b), (27c), and (28b).

(29) (a)

DET_1	-	(ADJ_1)	-	N_1	-	AND	-	DET_2	-	$(ADJ)_2$	-	N_2	
1		2		3		4		5		6		7	⇒
1		2		3		4		Ø		6		7	

where 1 = 5

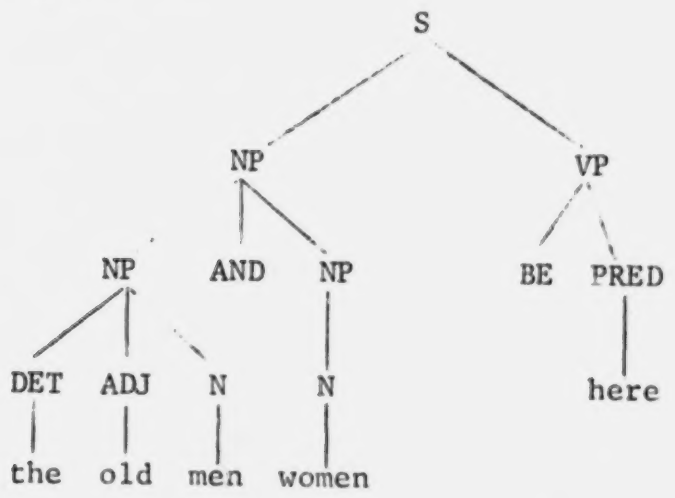
(b) ADJ_1 - N_1 - AND - ADJ_2 - N_2

1 2 3 4 5 ⇒
1 2 3 Ø 5

where 1 = 4

When the rules of (29) are applied to (27a) and (28a) to yield (27c) and (28b), respectively, the derived structure is in both cases that which is represented in (29c).

(29c)



Thus although the sentence 'The old men and women are here' is ambiguous (with readings (27) and (28)), the two readings have an identical surface structure. This in itself is not necessarily a defect, since it is known that other ambiguous sentences have this property (e.g., 'When did John promise to call?'). It might seem undesirable here, however, since we shall see in the next section that other ambiguities of a similar nature are indicated in the surface structure.

All coordinate constructions then can be accounted for by two types of conjunction processes: (1) The schema for coordination reduces separate sentences to a single sentence containing certain conjoined major constituents. We shall call the rules of this schema the primary conjunction rules. (2) A subsequent set of deletion rules may then optionally delete repeated elements of the

conjoined members. The deletion rules are of course transformational in nature. These rules will be referred to as secondary conjunction rules.

The secondary conjunction rules are not to be regarded merely as "fix-up" rules which handle those coordinate constructions not directly derivable from the primary conjunction rules. The deletion rules are well motivated. Two examples should suffice to illustrate this point.

Consider the conjoined nominal expression 'the old man and woman.' This construction, unlike the plural, is not ambiguous. It can only have the interpretation: 'the old couple.' Thus whereas the plural 'the old men and the women' can be further reduced to the old men and women, the singular 'the old man and the woman' (where the woman is not necessarily old) cannot be so reduced. That is, any string of the form DET_1 ADJ N_1 AND DET_2 N_2, where $DET_1 = DET_2$ can have DET_2 deleted only if N_2 is plural.[6] The deletion of the determiner is a special property of conjoined noun phrases containing plural nouns and is in no way attributable to any generalized principle of coordination of noun phrases. Thus, noun phrase coordination is independent of the subsequent deletion of articles with plural nouns.

Further justification for separate deletion rules can be found by examining coordination in different languages. For example, both French and English permit coordination of NP's: "The men and the women: les hommes et les femmes." In English the second determiner may be deleted. French, on the other hand, does not generally permit article deletion. Thus, whereas both languages permit coordination of the same type of grammatical categories, any further possible reduction processes within the conjoined constituents are not necessarily similar.

[6] If conjoined NP's both dominate the constituents DET ADJ N, then the DET (as well as the ADJ) can be deleted in the singular: The old man and the old woman → The old man and old woman → The old man and woman.

One could then think of the primary conjunction rules, which reduce conjoined sentences to a single sentence, as universal processes. It does not seem totally unreasonable to suppose that all languages have coordination of subjects, predicates, certain adverbials, etc. The principle for coordination (19) could then be stated independent of any particular language. The secondary conjunction rules, on the other hand, are language dependent; they account for alternants exhibited in a particular coordinate construction. These rules are optional transformational rules, and like other optional transformations within the syntactic component their effect is to produce stylistic variants.

6. EXAMPLES OF PRIMARY CONJUNCTION RULES[7]

We have shown that the primary conjunction rules should not yield conjoined N's but only conjoined NP's. This would seem to suggest that those constituents which are conjoinable are major grammatical categories which are not also lexical categories.[8] Thus, constituents such as N, V, ADJ will not be conjoined; instead, NP, VP, AP or other higher level constituents will enter into conjoined expressions. It is this additional restriction which is needed to characterize precisely the notion "constituent of like type." The principle of primary conjunction can then be stated as follows: Two (or more) sentences can be conjoined into a single sentence if their total tree structure is identical except for the structure dominated by one grammatical node; this node must be a major grammatical category which is not also a lexical category.

[7] Some secondary conjunction rules are to be found in the Appendix.

[8] "A category that appears on the left in a lexical rule we shall call a lexical category; a lexical category or a category that dominates a string... X..., where X is a lexical category, we shall call a major category... the categories N, V, and M are lexical categories, and all categories except Det (and possibly M and Aux) are major categories." N. Chomsky, Aspects of the Theory of Syntax, M. I. T. Press, 1965. p. 74.

In this section will be examined some of the constituents from separate sentences which can be conjoined (compounded) within a single sentence. Since the requirement is that all tree structure be identical except for the structure dominated by one grammatical node, sentences cannot be reduced if they have no tree structure in common (including lexical items) or if they have all their tree structure in common. Hence (30) and (31) are automatically excluded from primary conjunction.

(30) Mary goes to school during the day and John works at night.

(31) John works at night and John works at night.

Subject Position

We have already shown that the subjects of independent sentences can be conjoined if they have identical predicates.

(32) (a) The men PRES BE here and
The women PRES BE here.

(b) The men and the women are here.

Predicate Phrase

(33) (a) The plane SG PAST land at 1800 hours and the plane SG PRES will leave at 2030 hours.

(b) The plane landed at 1800 hours and will leave at 2030 hours.

Since all nodes must be identical except for one, the rule which expands S must be:[9]

(34) S → (PRE) NP PDP

if predicate phrase conjunction is to be allowed. The subsequent phrase structure rules for developing PDP are:

[9]The phrase structure rules presented here are a modified version of the rules found in the MITRE JUNIOR grammar (S010 of the English Preprocessor Manual). In the phrase structure rules (34) and (35) PRE is expanded to markers for interrogative and negative sentences; PDP is predicate phrase, TIM is time adverbial, VB is verbal, and AGNT is agent; the latter triggers the passive transformation.

(35) PDP → VP (TIM)

$$VP \rightarrow \left\{ \begin{matrix} VB & (NP & (AGNT)) & (ADV) \\ AUX & BE & PRED & \end{matrix} \right\}$$

VB → AUX V

These rules should become clearer in the course of the derivations. Motivation for this phrase structure is given in Section 7.

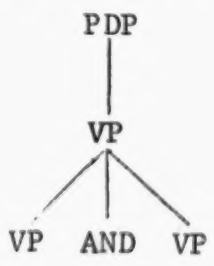
Verb Phrase

Verb phrases are conjoinable if all other constituents (including lexical items) are identical.

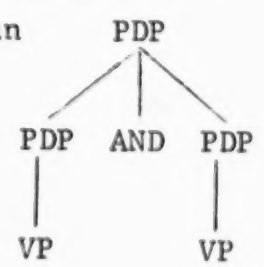
(36) (a) John PAST play the piano and John PAST sing.

(b) John played the piano and sang.

Note that in (36) 'PAST play the piano and PAST sing' can also be analyzed as a PDP. This is because PDP has been expanded to the single symbol VP. Therefore, we shall adopt the convention that for coordination purposes whenever a node X has Y as its only daughter and if both X and Y satisfy the principle for conjunction, then Y is to be selected as the conjoined constituent. This convention imposes the least amount of additional structure in the derived tree; e.g.,



is simpler than



It also implies that any further transformations which are applicable to coordinate constructions will not affect higher level constituents of the predicate phrase. The consequences of this assumption have not yet been explored.

Adverbial Position (TIME)

Adverbials are conjoinable if all other constituents (including lexical items) are identical.

(37) (a) John PAST sing on Monday and John PAST sing on Tuesday.

(b) John sang on Monday and on Tuesday.

Secondary conjunction rules could subsequently delete the repeated preposition, yielding

(38) John sang on Monday and Tuesday.

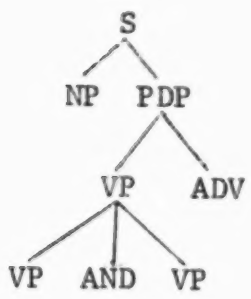
Sentence (39) is ambiguous since it can be derived from either (40a) or (40b). The derived structure is shown in (41a) (verb phrase conjunction) and (41b) predicate phrase conjunction.

(39) The plane landed and left on Monday.

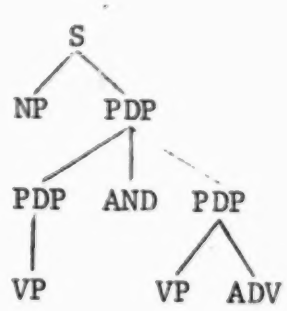
(40) (a) The plane SG PAST land on Monday and The plane SG PAST leave on Monday.

(b) The plane SG PAST land and The plane SG PAST leave on Monday.

(41a)



(41b)



Object Position

NP's are conjoinable in object position.

(42) (a) John PAST see Mary on Tuesday and John PAST see Helen on Tuesday.

(b) John saw Mary and Helen on Tuesday.

If both the NP and the adverbial are different as in (43), only PDP conjunction is possible since 'Mary on Tuesday' is not a constituent.

(43) John saw Mary on Tuesday and Helen on Thursday.

(43) has (44) as its source.

(44) John PAST see Mary on Tuesday and John PAST see Helen on Thursday.

Primary Conjunction yields (45).

(45) John saw Mary on Tuesday and saw Helen on Thursday.

(43) can be derived from (45) by secondary conjunction rules.

VB Conjunction

(46) (a) John PAST catch fish and John PAST eat fish.

(b) John caught and ate fish.

Note that (47) is ambiguous in that it is a case of either VB conjunction (48a) or VP conjunction (48b).

(47) John read and wrote letters.

(48) (a) John PAST read letters and John PAST write letters.

(b) John PAST read and John PAST write letters.

In fact (49) is three ways ambiguous being derived from (50a) (VB conjunction), (50b) (VP conjunction), (50c) (PDP conjunction).

(49) John read and wrote letters on Tuesday.

(50) (a) John PAST read letters on Tuesday and John PAST write letters on Tuesday.

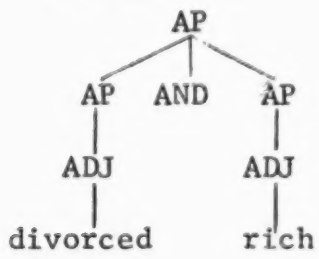
(b) John PAST read on Tuesday and John PAST write letters on Tuesday.

(c) John PAST read and John PAST write letters on Tuesday.

Adjective Conjunction

From the above examples it is seen that only major grammatical categories (which are not also lexical categories) can be conjoined. Such a principle would suggest that adjective conjunction (51) is really adjective phrase conjunction (52)

(51) John is divorced and rich.



The rules for adjective placement would have to take into consideration this structure.

Adjectival modifiers are derived from relative clauses; e.g., a man who is rich ⇒ a man rich ⇒ a rich man; a man who is red in the face ⇒ a man red in the face, but not *a red man in the face, *a red in the face man. Hence, an adjectival phrase (AP) cannot be preposed if it dominates an adjective complement.

7. PHRASE STRUCTURE RULES

(53) (a) $S \rightarrow (PRE)\ NP\ PDP$

(b) $PDP \rightarrow VP\ (TDM)$

(c) $VP \rightarrow \left\{ \begin{array}{l} VB\ (NP\ (AGNT))\ (ADV) \\ AUX\ BE\ PRED \end{array} \right\}$

(d) $VB \rightarrow AUX\ V$

The motivation for (53a) was given in Section 6. The phrase structure rules (53b) and (53c) allow sentences such as (54) to be conjoined, yielding (55), but automatically exclude conjoining (56) in such a way so as to yield the incorrect (57).

(54) *John PAST BE in Cambridge on Tuesday and John PAST see Mary on Tuesday.*

(55) *John was in Cambridge and saw Mary on Tuesday.*

(56) *John PAST BE in Cambridge and John PAST study in Cambridge.*

(57) **John was and studied in Cambridge.*

(56) can only yield (58) since in (56) the first occurrence of 'in Cambridge' is a PRED whereas the second occurrence of it is an ADV. Therefore, if the component sentences of (56) are to differ by the structure dominated by one grammatical node, that node must be VP.

(58) John was in Cambridge and studied in Cambridge.

This may then be pronominalized (see section 10); e.g. John was in Cambridge and studied there. Note that (55) is ambiguous. When derived from (54) it is a case of VP conjunction. However, (55) can also be derived from (59), PDP conjunction.

(59) John PAST BE in Cambridge and John Past see Mary on Tuesday.

The Aux Problem

Rule (53d)

$$VB \rightarrow AUX\ V$$

is needed due to the peculiar behavior of AUX.

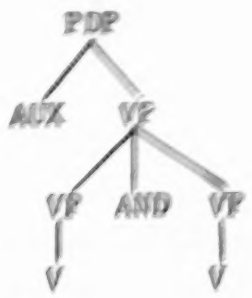
If AUX does not "hug" V, one would derive incorrect structures. Assume the rules (60), typical of phrase structure rules found in the literature.

(60) (a) PDP → AUX VP (TIM)

(b) $VP \rightarrow \left\{ \begin{matrix} V & (NP\ (AGNT)) & (ADV) \\ BE & PRED \end{matrix} \right\}$

With VP conjunction we might obtain

(61)



Then (62a,b) would yield incorrectly (63a,b), respectively.

(62) (a) John PAST sing and John PAST dance.

(b) John PRES BE ING sing and John PRES BE ING dance.

(63) (a) *John sang and dance.
(b) *John is singing and dance.

With the phrase structure rule (53d), only VB can be conjoined (not V since it is a lexical category). Then (62a,b) can only yield (64a,b).

(64) (a) John sang and danced.
(b) John is singing and is dancing.

Secondary conjunction rules may then delete the repeated is *in (64b) yielding.*

(65) John is singing and dancing.

A comparison between 64(a) and (b) shows that the secondary conjunction rules apply to words and not to inflectional morphemes.

Furthermore, the behavior of AUX provides additional motivation for the two types of conjunction rules. Consider sentences such as:

(66) The plane SG PAST can HAVE EN BE ING land at 1800 hours and
The plane SG PAST can HAVE EN BE ING leave at 1900 hours.

Primary conjunction applied to (66) yielós (67) (PDP conjunction).

(67) The plane could have been landing at 1800 hours and could have been leaving at 1900 hours.

Secondary conjunction rules then permit the derivation of all the intermediate reduced sentences.

(68) (a) The plane could have been landing at 1800 hours and have been leaving at 1900 hours.
(b) The plane could have been landing at 1800 hours and been leaving at 1900 hours.
(c) The plane could have been landing at 1800 hours and leaving at 1900 hours.

VB conjunction

If AUX V is always dominated by VB then 'sang and danced' and 'sang and will dance' are treated alike for conjunction purposes (i.e., both constructions are instances of VB conjunction). In fact, all the sentences of (69) illustrate VB conjunction.

(69) (a) John sings and dances.
(b) John sang and will dance.
(c) John can sing and will dance.
(d) John could have sung and should be dancing.

Note that AUX's cannot be conjoined. If such sentences as (70) are possible, they would be derived by means of secondary conjunction rules from their stressed counterparts (71).

(70) John can and does sing.
(71) John can sing and does sing.

8. INTERROGATIVES

Conjoined predicates (PDP, VP, or VB) can be questioned only if their AUX's are identical. Thus, only (72a,b) can be questioned in (73).

(72) (a) John sings and dances.
(b) John can sing and can dance.
(c) John sings and danced.
(d) John sings and can dance.
(e) John can sing and should dance.
(73) (a) Does John sing and dance?
(b) Can John sing and dance?
(c) *Does John sing and did dance?
(d) *Does John sing and can dance?
(e) *Can John sing and should dance?

Furthermore, not only must the AUX's be identical in conjoined questions but the AUXA (that part of the AUX which includes TNS and a following M, BE, or HAVE if present) can appear only once at the beginning of the conjoined construction. Thus, such sentences as (74) are not permitted.

(74) (a) *Does John sing and does dance?
(b) *Does John sing and dances?
(c) *Is John singing and is dancing?
(d) *Can John sing and can dance?

The problem here is that conjoined declaratives may show AUXA in all of the conjoined constituents (75a) and must show AUXA if it dominates only TNS (75b).

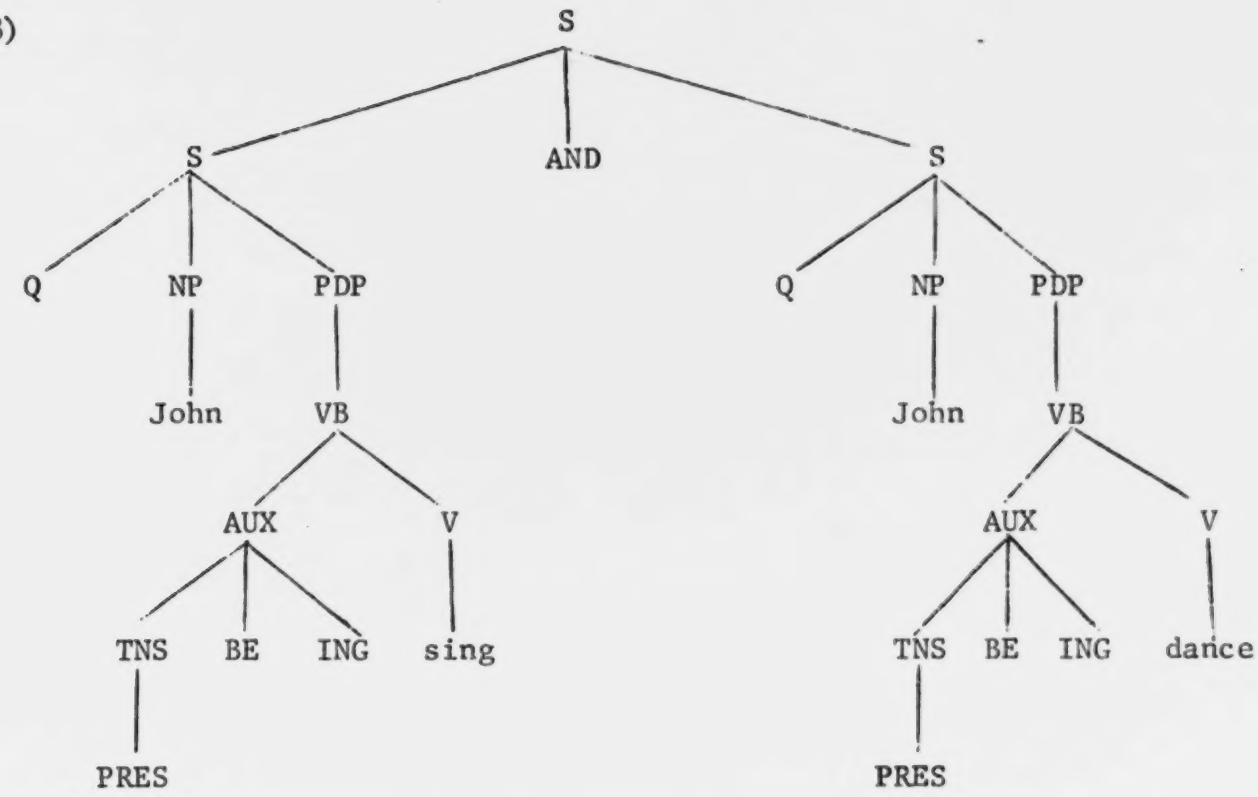
(75) (a) John is singing and is dancing.
(b) John sings and dances.

In the interrogative, TNS (plus M, BE, or HAVE), i.e., the AUXA, may not be repeated.

(76) (a) Is John singing and dancing?
(b) Does John sing and dance?

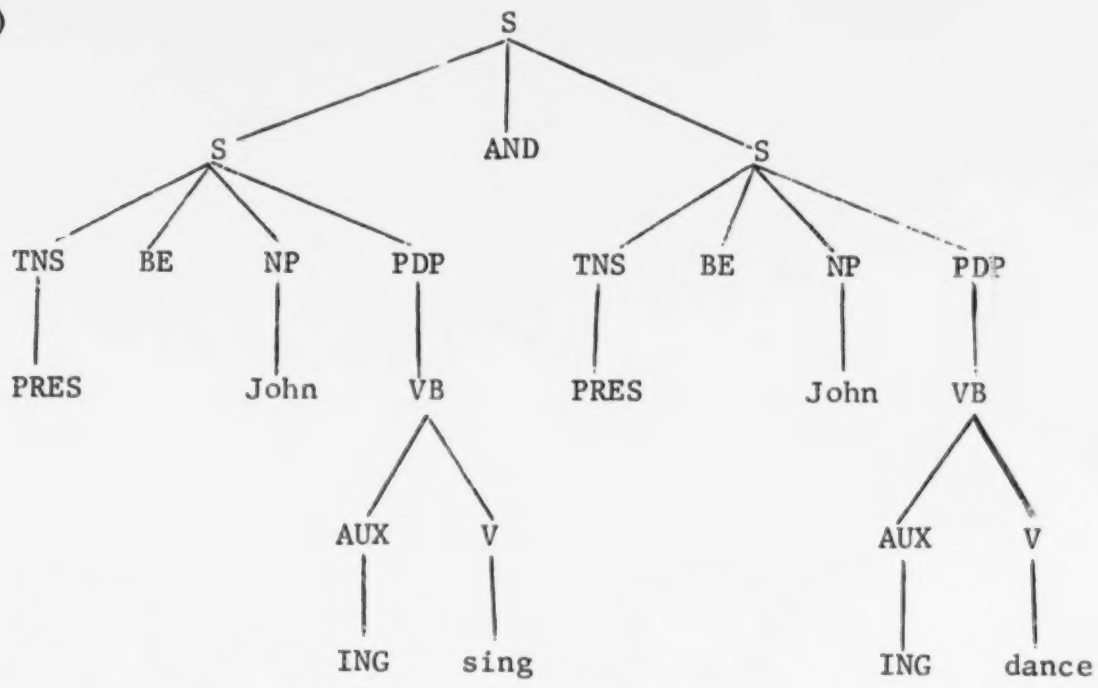
The correct set of sentences can be derived if we postulate the phrase structure rule (53d) and allow primary conjunction to take place after the interrogative transformation, which shifts AUXA to the head of the sentence. We shall derive (76a). Only the significant nodes are reproduced.

(78)



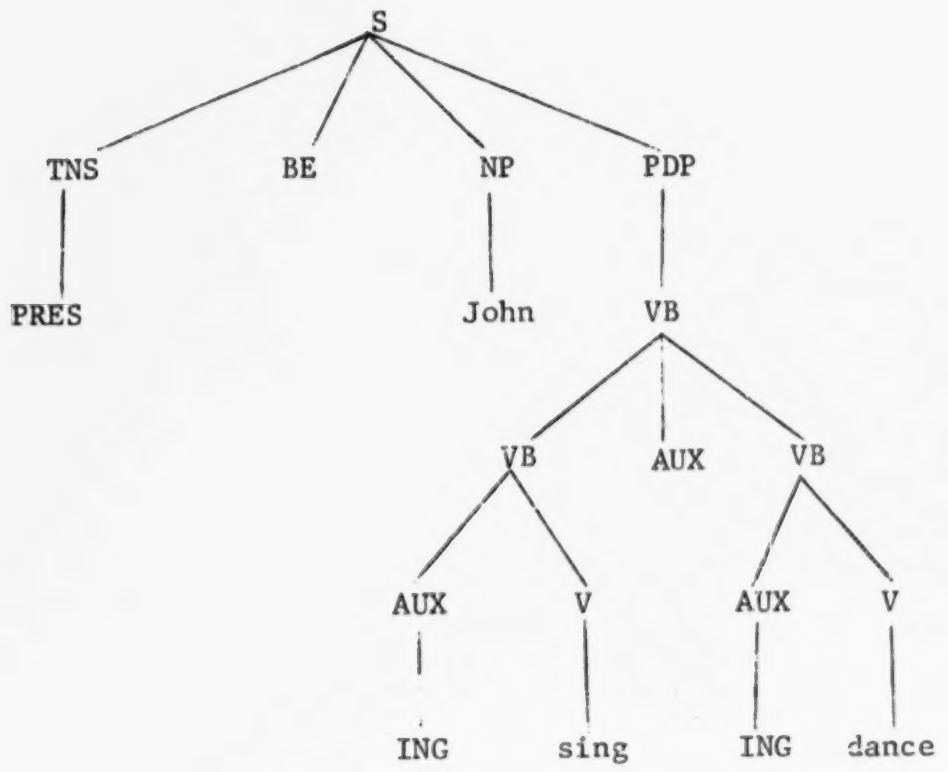
The interrogative transformation replaces the interrogative marker Q by AUXA (i.e., TNS + BE here).

(79)



Conjunction then applies; since all nodes are identical except VB, the result is (80).

(80)

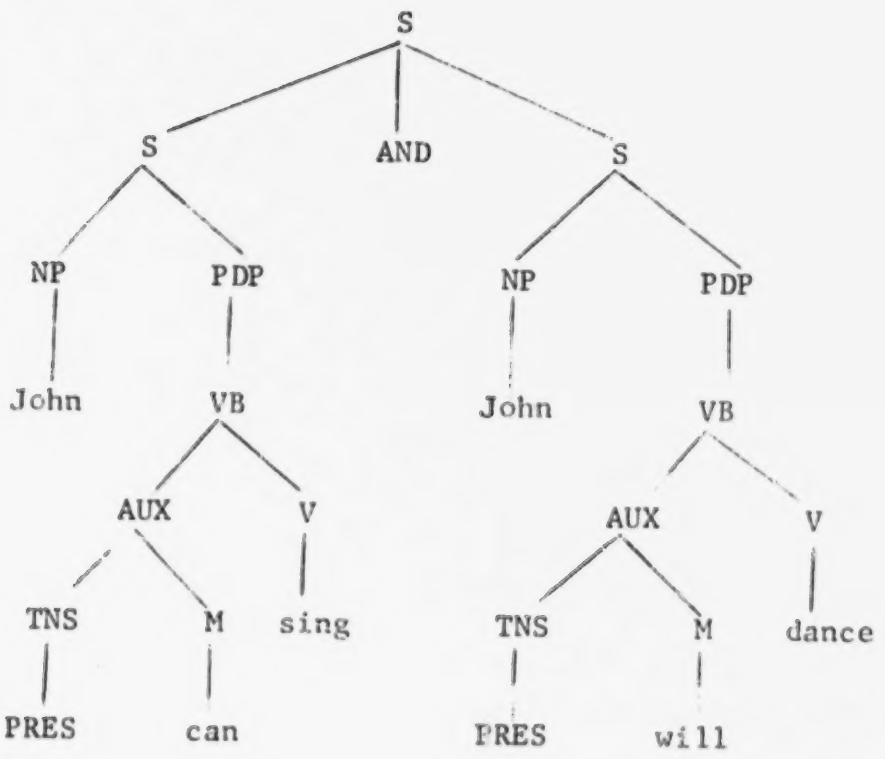


The transformation for affix placement will then produce the correct derived sentence.

The declarative sentences will have the same tree structure as (78) except there is no Q present; therefore, the transformation for the interrogative will not apply; AUXA will stay in the VP or VB and will not be collapsed by the conjunction rules. Then the full AUXA will emerge in the declarative.

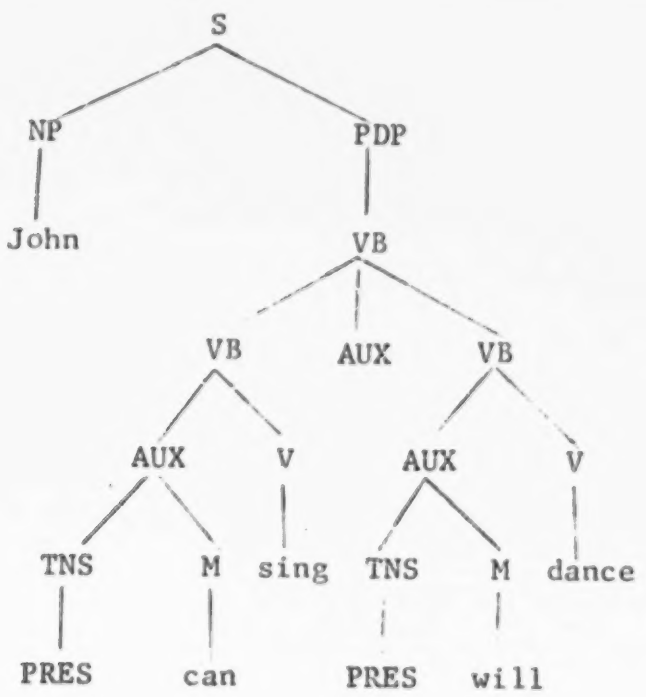
Whenever both the AUXA and the V are different, VB conjunction still occurs in the declarative.

(81)



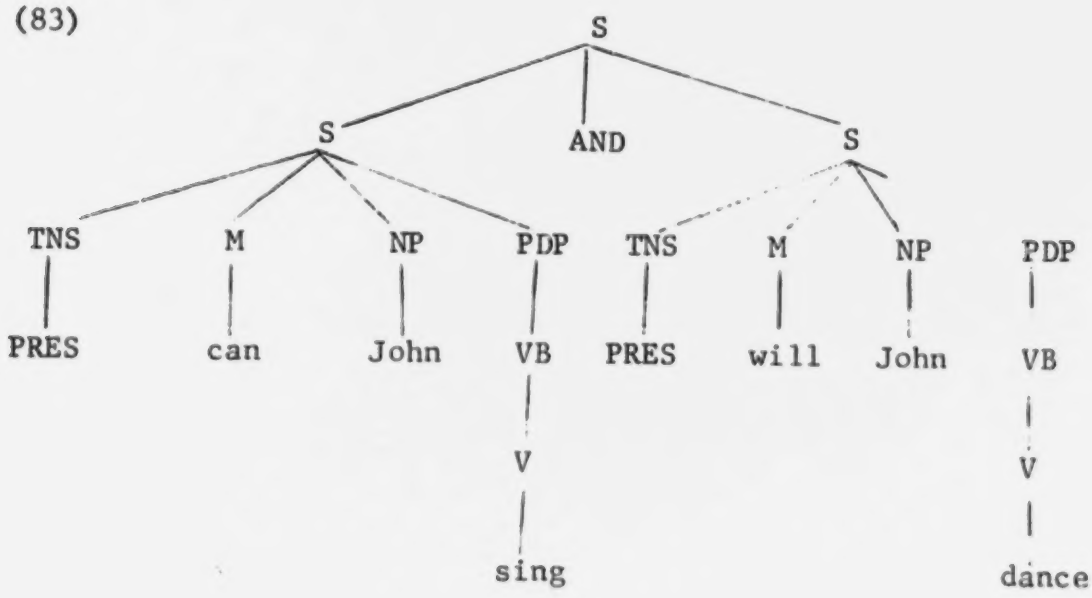
All nodes are identical except for VB; hence conjunction takes place.

(82)



If the AUXA's are not identical but a Q is present in each conjoined sentence, the interrogative transformation will of course apply and AUXA will be pulled out of the VB.

(83)



But now VB conjunction fails since the remaining nodes are no longer identical (i.e., the M are different). Therefore, we can only get conjunction of sentences.

9. THE OPTIONALITY OF PRIMARY CONJUNCTION

The primary conjunction rules are optional and can be applied to any single major grammatical node which satisfies the principle for conjunction. If the primary conjunction rules are not applied at a given level, the rules for pro-forms are applied obligatorily to certain repeated constituents. Thus, given the conjoined sentence (84), primary conjunction applied at the level of the VB yields (85).

(84) John PAST catch fish on Tuesday AND
John PAST eat fish on Tuesday.

(85) John caught and ate fish on Tuesday.

If primary conjunction is applied to (84) at the level of the VP, the result is (86).

(86) John caught fish and ate fish on Tuesday.

The repeated element (i.e. fish) is then obligatorily pronominalized, yielding:

(87) John caught fish and ate them on Tuesday.

If primary conjunction is applied to (84) at the level of the PDP, the result is (88).

(88) John caught fish on Tuesday and ate fish on Tuesday.

The repeated entities are replaced by the appropriate pro-forms.

(89) John caught fish on Tuesday and ate them then.

If primary conjunction is not applied to (84) then all the repeated entities are replaced by the appropriate pro-forms yielding:

(90) John caught fish on Tuesday and he ate them then.

10. SUBJECT-PREDICATE AGREEMENT

When the subject of a sentence is a conjoined nominal, the verb (as well as any predicate nominal) is plural.

(91) John and Mary are linguists.

This would suggest that the transformational rules for verb agreement (and predicate agreement) occur after primary conjunction. Now, within most transformational accounts of English, verb agreement takes place before the interrogative transformation--that is, the rule for verb agreement has the verb take on the number of the preceding N. Only after number has been attached to the verb can interrogative inversion take place. Thus, it would seem then that the order of transformations should be: primary conjunction, verb agreement, interrogative. Yet it has been shown in section 8 that due to the special

distribution of AUX within conjoined verbals, the interrogative transformation has to precede primary conjunction. If primary conjunction occurs after the interrogative transformation then verb agreement will fail to take place since the subject NP is no longer preverbal. This means either that our treatment of conjoined interrogatives is incorrect or else that the present formulation for verb agreement is not adequate.

If our treatment of conjoined interrogative constructions is invalid, then the present rules for verb agreement and their particular ordering should be appropriate. However, although the rules for verb agreement will work when the conjunction is <u>and</u> they will not describe those styles in which the declarative and interrogative forms of the same sentence may show differing verb number whenever the coordinate conjunction is <u>or</u>; i.e., the verb agrees with the member of the conjoined nominal which is closer to the verb.

(92) (a) The teacher or the students are right.
(b) Is the teacher or the students right?

Examples such as (92), although somewhat marginal, suggest that agreement appears relatively late in the grammar; or perhaps special agreement rules operate with compound subjects and are independent of the agreement rules for simple subjects. This possibility is not entirely unlikely for we have already demonstrated that coordinate sentences in many respects are unlike other sentence types.

Alternatively, in the deep structure (or perhaps after a subset of the transformations) one could mark the N which immediately precedes the V. Note that such marking has to be done anyway for the personal pronouns in order to obtain the correct case forms. This proposal would simply extend the case marking to N's as well. Then the agreement rule would require the verb to agree with the marked noun or pronoun (or in the case of a compound

subject, with specifically marked nouns or pronouns). Since the subject has been overtly marked, the actual order of the elements of the sentence need no longer be crucial when the agreement rule is applied. After application of agreement the case marker on nouns would be deleted. The consequences of this formulation are rather interesting. In some sense English nouns are still part of an inflectional system encompassing number as well as several cases. Although at the surface level the subject-object distinction, for example, has been obliterated in nouns, at a deeper level of analysis this distinction is still of functional importance. Is subject-verb agreement in Modern English not simply a vestige of what was a richer system in a previous stage of our language?

11. AN ADDITIONAL CONSTRAINT ON PRIMARY CONJUNCTION

The rule for primary conjunction was formulated as follows: Two (or more) sentences can be conjoined into a single sentence if their total tree structure is identical except for the structure dominated by one grammatical node; this node must be a major grammatical category which is not also a lexical category. However, this formulation would not exclude the sentences of (93) from being combined into the coordinate structures exhibited in (94).

(93) (a) The boy PRES BE here AND a boy PRES BE here.
(b) The boy PRES BE here AND the boys PRES BE here.
(c) The old man PRES BE here AND the man PRES BE here.
(d) John PRES sing AND John PAST sing.

(94) (a) *The boy and a boy are here.
(b) *The boy and the boys are here.
(c) *The old man and the man are here.
(d) *John sings and sang.

Compare these with:

(95) (a) The boy PRES BE here AND a girl PRES BE here.
(b) The boy PRES BE here AND the girls PRES BE here.

(c) The old man PRES BE here and the woman PRES BE here.
(d) John PRES can sing AND John PRES will sing.

(96) (a) The boy and a girl are here.
(b) The boy and the girls are here.
(c) The old man and the woman are here.
(d) John can sing and will sing.

Note that the major grammatical categories which can be conjoined always dominate one or more lexical categories and may dominate non-lexical entities (e.g., articles, suffixes) as well. In order for primary conjunction to take place each of the conjoinable members must exhibit <u>at least one difference</u> (i.e., a different lexical item) <u>within the same lexical category</u>. For example, in (93a) the two noun phrases differ only in the article. Since the article is not a lexical category, primary conjunction cannot take place. In (95a) although the noun phrases differ in the article they also differ in the noun. Therefore, primary conjunction takes place. Thus, primary conjunction requires that there be different lexical items from the same lexical category. However, the lexical categories themselves are not directly conjoinable but rather it is the larger units which dominate lexical categories, which ultimately are conjoined.

12. CONJUNCTION AND THE ENGLISH PREPROCESSOR

The schema for primary conjunction has been programmed for the computer (see Appendix B). The program makes use of a list of major constituents which are conjoinable (i.e., NP, PDP, VP, VB, LOC, AP, TIM). If a form can be analyzed in more than one way, minimal analysis is taken; i.e., if a VP node satisfies the conditions on conjunction, then the PDP node dominating it will also satisfy those conditions, but conjunction will apply to the lower node, VP. Surface grammar rules and reversal rules are presently in preparation.

A small number of sentences have been tested on the computer making use of a simplified version of the MITRE JUNIOR Grammar (see S010 of the English Preprocessor Manual). Rules 2 and 4 in the JUNIOR phrase structure were changed to incorporate the phrase structure presented in section 8. Other phrase structure rules (i.e., those pertaining to the expansion of the determiner system) were eliminated so as not to introduce extraneous complications in the initial testing stage. Transformational rules for secondary conjunction were added to the transformational component of the grammar. (A list of phrase structure and transformational rules is given in Appendix A).

To date we have tested conjoined NP's in subject, object, and predicate positions. The deletion rules pertaining to NP's for secondary conjunction have been debugged. We have also tested conjoined PDP's, VP's, and VB's. The rules for auxiliary reduction both in declarative and interrogative sentence types are operating correctly. The rules for subject-verb agreement (with compound subject) have been tested and seem adequate. However, we have not completely debugged the rules which operate on conjoined relatives and adjectives.

We have tested the primary conjunction rule on as many as four conjoined sentences. We have also made provision for the reapplication of the primary conjunction rule, and have successfully derived the sentence: 'Sandy and Mary like iguanas and snakes' in two steps from the four underlying sentences.

(a) Sandy PRES like snakes AND
(b) Sandy PRES like iguanas AND
(c) Mary PRES like snakes AND
(d) Mary PRES like iguanas.

The primary conjunction rule applies separately to (a,b) and to (c,d,), yielding:

(e) Sandy PRES like snakes and iguanas AND

(f) Mary PRES like snakes and iguanas.

The primary conjunction rule is then re-applied to (e,f), yielding:

(g) Sandy and Mary PRES like snakes and iguanas.

PRECEEDING PAGE BLANK NOT FILMED

APPENDIX A

The Test Grammar

PHRASE STRUCTURE RULES

1. SS → #S#

2. S → { #S# ({ AND / OR } #S#)* / (PRE) NP PDP }

3. PRE → Q

4. PDP → VP (TIM)

5. VP → { VB (NP (AGNT)) (ADV) / AUX BE PRED }

6. VB → AUX V

7. TIM → TM

8. ADV → LOC

9. V → { VTR/ ___NP / VINT }

10. → { AP / NP / (LIM) (LOC) }

11. AP → ADJ

12. AGNT → BY

13. LOC → { ({ AT / IN }) NPP / HERE }

14. TM → AT NPI

15. NPP → DET N NU

16. NPI → { DET TIME NU / INT HOURS }

17. NP → DET N NU (SS)

18. N → { NPR / NCM }

19. NCM → { NCT / NMS }

20. DET → { NL/ART / ___NPR/PSAR }

21. NU → { A / THE / DEM }

22. NU → SG/ { NPR / NMS / TIME } _____

23. NU → { SG / PL }

24. INT → PLINT

25. DEM → (WH) THAT NBR

26. PSAR → ADM

27. AUX → AUXA (M) (HAVE EN) (BE ING)

28. AUXA → TNS

29. TNS → { PRES / PST }

TRANSFORMATIONS

Many of the transformational rules are identical to those in the JUNIOR Grammar. These rules then will not be stated nor will examples of them generally be given since they are to be found in SR-132, English Preprocessor Manual. However, the name of the rule with a brief description will still appear in the ordered set of transformations.

1. TPAS

   The Passive Transformation has to apply before Conjunction in order to derive conjoined nominals, such as: The Tigers beat the Yankees and were beaten by the Red Sox.

3. TAX1

4. TAX2

22. TAX3

```
($SKIP  Q)   (NP)   (AUXA)
     1         2      3

(3)  ADRIS  2

ERASE  3
```

TAX1, TAX2, and TAX3 operate on the auxiliary. They do not rearrange the order of constituents in the AUX but rather change the tree structure in preparation for the Negative and Interrogative Transformations. Thus, Phrase Structure Rules 27, 28, and 29 expand AUX into AUXA (M) (HAVE EN) (BE ING); AUXA into TNS; and TNS into PRES or PST. Initially, then, AUXA contains only TNS; transformational rules TAX1 and TAX2 incorporate into AUXA a following M, HAVE, or BE (if present) yielding as a single constituent the $\text{TNS}\left(\left\{\begin{array}{l}\text{M}\\ \text{HAVE}\\ \text{BE}\end{array}\right\}\right)$ needed for interrogative and negative constructions. After TAX1 and TAX2 the AUXA is still dominated by AUX (see The English Preprocessor Manual, S003, p. 55). TAX3 attaches AUXA as a right sister to the preceding NP (i.e., AUXA becomes immediately dominated by S), whenever the interrogative marker Q is present. This transformation is

needed so that AUXA will be reduced appropriately in coordinate constructions (see Section 8 for discussion.) This treatment of the auxiliary in the MITRE grammar is not the same as the process outlined in Section 8 since there are no TAX transformations there. The major difference in approach is that TAX3 is a device for allowing the proper auxiliary conjunction to take place before the interrogative transformation is applied. This order is needed since in the MITRE grammar subject-verb agreement with conjoined nominals must occur after conjunction and before interrogation.

Primary Conjunction Rule

TC (cyclical, embedding)

($SKIP #)	(S)	(#)	(AND OR)	(#)	(S $RES 10)	(#)
1	2	3	4	5	6	7

$RES 10: neq 2

NOTE: The operation statements are performed on the nodes as renumbered following resetting by the conjunction subroutine. For discussion see Appendix B: Programming of Conjunction.

The operation statements are:

1 SUBST 2

2 ADLAD 1

4 ADLAD 1

3 ADLAD 1

ERASE 2,3,4,5,6,7

2. TNU2

23. TNU3

($SKIP NP $RES 11)	(AND)	(NP $RES 11)	(TNS)	(SG)
1	2	3	4	5

(PL) SUBST 5 $RES 11: ndom NP

TNU2 causes the verb to take the same number as the subject. TNU3 provides for a plural verb whenever there is a compound subject whose members are conjoined by the conjunction AND. TNU2 must apply after conjunction so that subject NPs which differ in number may be conjoined; e.g., The girl PL PRES BE here and the boy SG PRES BE here. Note that when the conjunction rules apply the verb phrases do not yet have number. Since they are identical at this point coordination is possible; e.g., The girl PL and the boy SG PRES BE here ⇒ The girl PL and the boy SG PRES SG BE here (TNU2) ⇒ The girl PL and the boy SG PRES PL BE here (TNU3) ⇒ The girls and the boy are here.

5. TTH

TTH accounts for sentences with introductory there; e.g., There are a linguist and a programmer in the room, which is derived from A linguist SG and a programmer SG PRES PL BE LDM in the room, where LDM is a marker which triggers TTH.

6. TFXO

Rejection of all strings containing LDM to which TTH failed to apply--i.e., all strings containing LDM in which the subject NP does not dominate the indefinite article.

8. TAPL

Deletion of the indefinite article before plural nouns.

9. TAMS

Deletion of the indefinite article before mass nouns.

11. TNBR

Number marking for demonstratives.

12. TWH1

Movement of a noun phrase containing WH to the front of a non-interrogative sentence (in particular, to the front of an embedded relative sentence).

13. TREL

Embedding of relative clauses.

14. TADJ

The preposing of adjectives before a noun.

17. TQ

Movement of AUXA to the beginning of the sentence in questions.

18. TFX3

Rejection of sentences still containing WH. This situation could only arise if a sentence originally contained more than one WH.

19. TERA

Erasure of various markers which were used to trigger transformations.

25. TSEN

```
($SKIP  #)  (AND  OR)  (#)
   1            2      3
```

ERASE 1, 3

TSEN deletes the word boundary on either side of the conjunction in those compound sentences which have not undergone primary conjunction:

Example: # THE plane PAST arrive # AND # THE train PAST leave # ⇒

THE plane PAST arrive AND THE train PAST leave #.

20. TFIN

Any string with an internal # is completely erased. This situation arises, for example, when identity conditions for forming relative clauses are not met.

Secondary Conjunction Rules

26. DART

```
($SKIP  ART)  (ADJ)  (N)  (NU)  (AND  OR)
     1          2     3    4       5

(ART  $RES  12)  (ADJ  $RES  13)  (N)  (NU  $RES  15)
     6                7            8         9
```

ERASE 6

$RES 12: eq 1
$RES 13: eq 2
$RES 15: eq 4

Article deletion

Example: The old man and the old woman ⇒

The old man and old woman

27. DARP

($SKIP ART) (ADJ) (N) (PL) (AND OR) (ART $RES 12) (N) (PL)
1 2 3 4 5 6 7 8

ERASE 6

$RES 12: eq 1

Article deletion with plural noun

Example: The old men and the women ⇒

The old men and women

28. DADJ

($SKIP ADJ) (N) (NU) (AND OR) (ADJ $RES 12) (N) (NU $RES 14)
1 2 3 4 5 6 7

ERASE 5

$RES 12: eq 1
$RES 14: eq 3

Adjective deletion

Example: An old man and old woman (see DART) ⇒

An old man and woman

29. ARTD

($SKIP ART) (N) (NU) (AND OR) (ART $RES 12) (N) (NU $RES 14)
1 2 3 4 5 6 7

ERASE 5

$RES 12: eq 1
$RES 14: eq 3

Article deletion (no ADJ present)

Example: A man or a woman ⇒

A man or woman

30. DELN

(\$SKIP ART) (ADJ) (N) (NU) (AND OR) (ART \$RES 12)
1 2 3 4 5 6

(ADJ) (N \$RES 14) (NU \$RES 15)
7 8 9

ERASE 3,4

\$RES 12: eq 1
\$RES 14: eq 3
\$RES 15: eq 4

Noun deletion

Example: A green car or a red car ⇒

A green or a red car

31. DAPL

(\$SKIP ART) (ADJ) (AND OR) (ART \$RES 12) (ADJ) (N) (PL)
1 2 3 4 5 6 7

ERASE 4

\$RES 12: eq 1

Article deletion with plural noun

Example: The green and the red cars (see DELN) ⇒

The green and red cars

32. DNPL

(\$SKIP ADJ) (N) (PL) (AND OR) (ADJ) (N \$RES 13) (PL)
1 2 3 4 5 6 7

ERASE 2, 3

\$RES 13: eq 2

Deletion of noun

Example: The green cars and red cars (see DART) ⇒

The green and red cars

Note: DNPL and DAPL can produce identical surface structures.

33. DVB1

($SKIP VB) (NP AGNT) ($NIL ADV) (TIM) (AND OR)
1 2 3 4 5

(VB $RES 12) (NP AGNT) ($NIL ADV) (TIM)
6 7 8 9

ERASE 6 $RES 12: eq 1

Deletion of verb in predicate containing time adverbial

Example: John saw Mary at 8 O'clock and saw Helen at 2 O'clock ⇒

John saw Mary at 8 O'clock and Helen at 2 O'clock

34. DVB2

($SKIP VB) (NP AGNT) (ADV) (AND OR) (VB $RES 12)
1 2 3 4 5

(NP AGNT) (ADV)
6 7

ERASE 5 $RES 12: eq 1

Deletion of verb in predicate containing adverb

Example: John saw Mary in Boston and saw Helen in Detroit ⇒

John saw Mary in Boston and Helen in Detroit

In the given phrase structure grammar, ADV can only dominate LOC. With a larger phrase structure base, this transformation would have to be restricted so as to exclude sentences like:

* John saw Mary in Boston and Helen briefly

35. DAX1

($SKIP TNS) (NU) (M HAVE BE) ($NIL EN ING) (HAVE BE)
1 2 3 4 5

(EN ING) ($SKIP AND OR) (TNS $RES 12) (NU $RES 13)
6 7 8 9

(M HAVE BE $RES 14) ($NIL EN ING $RES 15)
10 11

(HAVE BE $RES 18) (EN ING $RES 19)
12 13

ERASE 8, 9, 10

$RES 12: eq 1
$RES 13: eq 2
$RES 14: eq 3
$RES 15: eq 4
$RES 18: eq 5
$RES 19: eq 6

Reduction of auxiliary (Part 1)

Examples: (a) PST SG can HAVE EN (BE ING) (BE EN) take and

PST SG can HAVE EN (BE ING) (BE EN) use ⇒

PST SG can HAVE EN (BE ING) (BE EN) take and

HAVE EN (BE ING) (BE EN) use.

(b) PST SG can BE ING (BE EN) take and

PST SG can BE ING (BE EN) use ⇒

PST SG can BE ING (BE EN) take and

BE ING (BE EN) use.

(c) PST SG can BE EN take and

PST SG can BE EN use ⇒

PST SG can BE EN take and

BE EN use

(d) PRES PL HAVE EN BE ING (BE EN) take and

PRES PL HAVE EN BE ING (BE EN) use ⇒

PRES PL HAVE EN BE ING (BE EN) take and

EN BE ING (BE EN) use.

(e) PRES PL HAVE EN BE EN take and

PRES PL HAVE EN BE EN use ⇒

PRES PL HAVE EN BE EN take and

EN BE EN use.

(f) PST PL BE ING BE EN take and

PST PL BE ING BE EN use ⇒

PST PL BE ING BE EN take an

ING BE EN use.

36. DAX2 (cyclical)

($SKIP HAVE BE) (EN ING) ($SKIP AND OR) ($NIL EN ING)
1 2 3 4

(HAVE BE $RES 12) (EN ING $RES 13)
5 6

ERASE 4, 5

$RES 12: eq 1
$RES 13: eq 2

Reduction of auxiliary (Part 2)

Examples: (a) could { HAVE EN / BE ING / BE EN } take and { HAVE EN / BE ING / BE EN } use ⇒

could { HAVE EN / BE ING / BE EN } take and { EN / ING } use

(b) could { HAVE EN / BE ING } BE EN take and { HAVE EN / BE ING } BE EN use ⇒

could { HAVE EN / BE ING } BE EN take and { EN / ING } BE EN use

Reapplication of DAX2 to above yields:

could { HAVE EN / BE ING } BE EN take and EN use.

(c) could HAVE EN BE ING BE EN take and

HAVE EN BE ING BE EN use ⇒

could HAVE EN BE ING BE EN take and

EN BE ING BE EN use

Reapplication of DAX2 to above yields:

could HAVE EN BE ING BE EN take and

ING BE EN use

Reapplication of DAX2 to above yields:

could HAVE EN BE ING BE EN take and EN use

(d) TNS NU { HAVE EN / BE ING } BE EN take and { EN / ING } BE EN use ⇒

TNS NU { HAVE EN / BE ING } BE EN take and EN use

(e) TNS NU HAVE EN BE ING take and EN BE ING use ⇒

TNS NU HAVE EN BE ING take and ING use

(f) TNS NU HAVE EN BE ING BE EN take and

EN BE ING BE EN use ⇒

TNS NU HAVE EN BE ING BE EN take and

ING BE EN use

Reapplication of DAX2 to the above yields:

TNS NU HAVE EN BE ING BE EN take and EN use

37. DAX3

```
(SKIP TNS)   (NU)   (M HAVE BE)   (NIL EN ING)   (V)   (SKIP AND OR)
     1         2         3              4          5          6

(TNS $RES 12)   (NU $RES 13)   (M HAVE BE $RES 14)   (NIL EN ING $RES 15)
      7               8                 9                    10

ERASE 7, 8, 9
```

$RES 12: eq 1
$RES 13: eq 2
$RES 14: eq 3
$RES 15: eq 4

Reduction of auxiliary (Part 3)

Examples: (a) TNS NU can take and TNS NU can use ⇒

TNS NU can take and use

(b) TNS NU { HAVE EN / BE ING / BE EN } take and TNS NU { HAVE EN / BE ING / BE EN } use ⇒

TNS NU { HAVE EN / BE ING / BE EN } take and { EN / ING } use

39. REJ

($SKIP ART) ($NIL ADJ) (N) (NU) (AND OR) (ART)
1 2 3 4 5 6

($NIL ADJ $RES 13) (N $RES 14)
7 8

ERASE Ø

$RES 13: eq 2
$RES 14: eq 3

Rejection of a sentence containing a conjoined NP where all constituents are identical except for article or number of noun or else both.

Examples: (a) The (young) man and a (young) man ⇒ Ø.

(b) The (young) man and the (young) men ⇒ Ø.

(c) A (young) man and the (young) men ⇒ Ø.

APPENDIX B

Programming of Conjunction

Joyce Friedman
Stanford University

1. STRATEGY

SYNN, an experimental analysis program for transformational grammars (WP-229), was designed to implement the MITRE analysis procedure for transformational grammars. The mechanism for programming conjunction was constructed to take advantage of routines already in the SYNN program. Minor revisions were made in the routine for applying forward transformations and in the analysis routine. Two new subroutines, CMARK, to mark levels of SS's, and CONJ, to search for conjoinable pairs of subtrees, were added. None of the changes affects the program as it now applies to the SENIOR grammar.

In the application routine (APPFXC) conjunction transformations were added as a separate cyclical group with code number 0. The pattern of cycling is the same as for the embedding transformations. Conjunction is performed first on the lowest SS's, then on the next lowest, and so on. (As currently coded, only three levels of SS's are considered.)

The analysis routine (XFAN) was modified only to allow reentry at a point where the search can be resumed. The failure (in the conjunction subroutine) to find a conjoinable pair is thus treated as an unsuccessful partial analysis and the analysis continues where it left off.

2. PRIMARY CONJUNCTION TRANSFORMATION

A primary conjunction transformation was added:

TC cyclical, embedding[1]

($SKIP +)	(S)	(+)	(AND OR)	(+)	(S $RES 10)	(+)
1	2	3	4	5	6	7

$RES 10: not eq 2

The instructions are not carried out until the numbers of the nodes satisfying the analysis have been reset by the conjunction subroutine. The instructions are:

---

[1]The symbol + as used in Appendix B is equivalent to the symbol #.

1 SUBST 2

2 ALADE 1

4 ALADE 1

3 ALADE 1

ERASE 5, 6, 7

This transformation is called embedding for technical reasons having to do with the analysis program. This does not necessarily have any linguistic implications.

3. CONJUNCTION SUBROUTINE

The conjunction subroutine CONJ is entered immediately after the analysis subroutine XFAN has found the analysis for TC. It determines which constituents, if any, can actually be conjoined. If none are found, XFAN is reentered to resume the search.

Since only "major constituents" can be conjoined, a list of them is built into the program. They are NP, PDP, VP, VB, AP, TIM, AND ADV. Of these all but PDP and VP are marked as immediately conjoinable. PDP or VP will be conjoined only if there are no elements below them which are conjoinable.

In the conjunction subroutine the elements to be conjoined are designated N3 and N4. These are initially set to zero, and are reset whenever a conjoinable pair is found. The search for conjoinable elements begins with the two symbols S found by the analysis routine. These are the initial values of N1 and N2. The search is shown by the flow chart in Figure 1.

4. INSTRUCTIONS FOR CONJUNCTION

In order to use the transformation format for conjunction, a somewhat unnatural use has been made of the instruction set for TC. The node numbers are originally set by XFAN to those of the analysis.

(+) (S) (+) (AND OR) (+) (S) (+)

If N3 is not already a compound constituent, the first three numbers of the analysis are reset by CONJ to yield the numbers of

(X) (N3) (N4) (AND OR) (+) (S) (+)

where X is a new node having the same symbol as N3 and N4. The instructions then yield a tree in which N3 has been replaced by

and all of (AND OR) (+) (S) (+) has been deleted. If N3 is already a compound constituent, the numbers are set to correspond to

(N3) ($NIL) (N4) (AND OR) (+) (S) (+)

In this case AND/OR and N4 are added as last daughters of N3, and the (AND OR) (+) (S) (+) deleted. The convention that an operation n XXXXX m is vacuous if either n or m is $NIL makes it possible to use the one set of instructions in both cases. The instructions are carried out by the XFDO subroutine as usual.

CONJ is entered from XFAN with node numbers for two sentence S's which are to be conjoined if possible. The analysis is (+) (S) (+) (AND OR) (+) (S $RES 1) (+), $RES 1: neq 2.
1 2 3 4 5 6 7

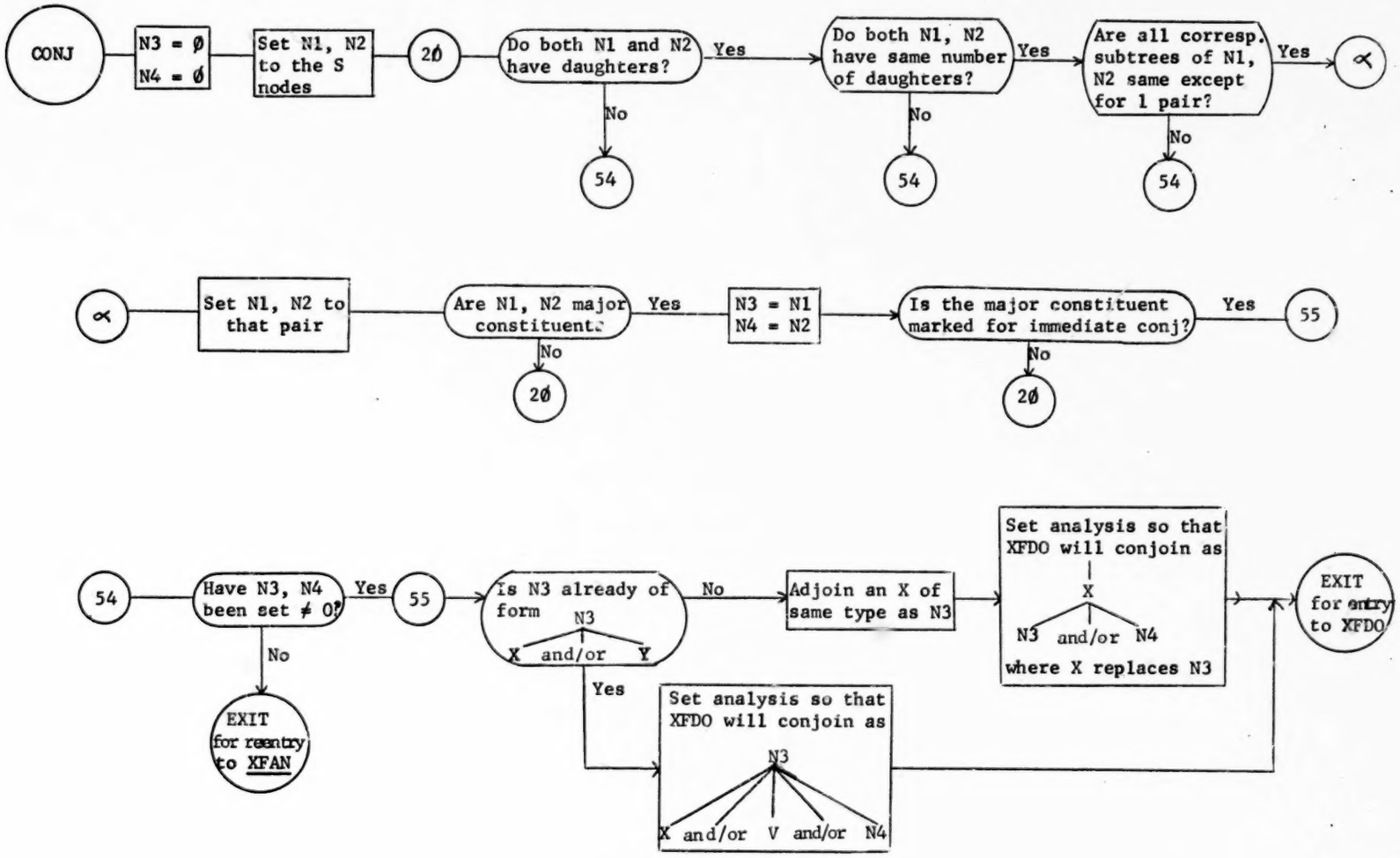


Figure 1. Conjunction Subroutine

APPENDIX C

References for the MITRE Grammar and Analysis Procedure

English Preprocessor Manual. SR-132, MITRE Corporation, 1964, 1965.

Zwicky, A. M., Friedman, J., Hall, B. C., and Walker, D. E. The MITRE Analysis Procedure for Transformational Grammars. AFIPS Conference Proceedings: 1965 Fall Joint Computer Conference 1965, 27, 317-326. Also MTP-9, MITRE Corporation, 1965.

Friedman, J. SYNN, An Experimental Analysis Program for Transformational Grammars. WP-229, MITRE Corporation, 1965.

Chapin, P. G., and Geis, M. L. The MITRE Grammar (January 1966). MTR-121, MITRE Corporation, 1966.